RANKING BASED ON TRIPLE COMPARISON

A Thesis

by

LIZI ZHANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Chao Tian
Committee Members,	Tie Liu
	Xiaoning Qian
	Anxiao Jiang
Head of Department,	Miroslav M Begovic

May 2020

Major Subject: Electrical Engineering

ABSTRACT

In this paper, we are interested in deducing the order of a set of items, under certain practical constraints (e.g., difficult to rank all of them at the same time, or having noise in the ranking process), only noisy partial orders on smaller subsets with a specific cardinal of the items can be obtained. For example, 10 cyclists are going to race with speed, but the track only allows 3 of them to compete simultaneously. How to get a full rank of them if the observing outcome will always be a partial ranking?

Generally speaking, how do we congregate these noisy partial ranking results into a complete ranking, and under what condition can we guarantee the resulting ranking to be accurate? These are the questions we seek to develop understanding in this work.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of Professor Tian Chao advisor, Professors Tie Liu and Xiaoning Qian of the Department of ECE and Professor Anxiao Jiang of the Department of Computer Science.

All work for the thesis was completed by the student, in collaboration with Wenjing Chen of the Department of Electrical and Computer Engineering Department.

Funding Sources

There are no outside funding contributions to acknowledge related to the research and compilation of this document.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

There are substantial papers discussed the problem of finding approximate rankings based on comparisons. Several papers (e.g., [1], [1], [2], [3]) introduced several ranking algorithm

Another papers [[4], [5], [6], [7], [8]] using parametric models such as BTL or PL and recovered the parameters associated to the individual item. The work is recently generalized to strong stochastic transitivity (SST) models, a more general class of models while leaving the question of whether or not these results can directly extend to tight bounds.

The works [[9], [10], [11], [12]] covered mixture models under the problem of recovery of the top k items. The [9] analyzes the performance of an algorithm that approximately ranks from pairwise comparisons. In [10] it is assumed that the preferences across individuals can be described by a low-dimensional modal. [11] has shown how the top-ranked items from pairwise comparisons can be resolved using a maximum entropy distribution technique.

Paper more recently [[13], [14], [15], [16]] has more connections to our paper. Wauthier et al. [13] addressed a weighted counting algorithm to recover approximate rankings. They consider recovery of an approximate ranking (under Kendall's tau and maximum displacement metrics) but do not provide results on exact recovery. Meanwhile, their bounds are quite loose: their results are tight only when there are a total of at least $\Theta(n^2)$comparisons, which is not a practical requirement.

The pair of papers [[14], [15]] by Rajkumar et al. consider ranking under several models and several metrics. In the part that is common with our setting, they show that the counting algorithm is consistent in terms of recovering the full ranking, which automatically implies consistency in exactly recovering the top k items. They obtain upper bounds on the sample complexity in terms of a separation threshold that is identical to a parameter Δ_k defined subsequently in this paper (see Section 3). However, as our analysis shows, their bounds are loose by at least an order of magnitude. They also assume a certain high-SNR condition on the probabilities, an assumption that is not imposed in our analysis.

Chen and Suh [16] proposed an algorithm called the Spectral MLE for the exact recovery of the top k items. They showed that, if the pairwise observations are assumed to draw according to the Bradley-Terry-Luce (BTL) parametric model [[17]], the Spectral MLE algorithm recovers the k items correctly with high probability under certain regularity conditions. Besides, they also show, via matching lower bounds, that their regularity conditions are tight up to constant factors. However, these guarantees are restricted to cases where data must be drawn from the BTL model.

In this paper, I will first shortly discuss the background of this problem (section 2.1), then provide a formal description of this problem (section 2.2), after that, a learning algorithm derived from the basic Copeland algorithm will be addressed.

Our main result of this project will be the analysis of the algorithm (section 3). We will study the optimality by proving a theorem which tells the upper bound of this algorithm under certain condition. Also, we will discuss the lower bound of this algorithm, as a converse part of the proof. more details will be covered in section 3.

In section 4, I will describe how to justify the theorem we proposed in section 3.

2. BACKGROUND AND PROBLEM STATEMENT

Our thesis mainly bases on the Nihar's 2016 paper: Simple, Robust and Optimal Ranking from Pairwise Comparisons. In this article, Nihar declared that the Copeland algorithm is optimal given pairwise comparison data by proving one of the three theorems he proposed. Inspired by his work, we will firstly generalize the Copeland algorithm to handle when we are given partial ranking data instead of a series outcome of the comparison.

2.1 Problem Statement

We consider the following scenario: there are a total of n items indexed in the set $[n] = 1, 2, ..., n$, and the samples on the partial rankings are collected in r rounds. In round $l = 1, 2, ..., r$, given a subset of n items whose cardinal is 3, say $\mathcal{A} \subseteq [n]$, where $|\mathcal{A}| = 3$, whose noisy ranking is observed with probability p, and not observed with probability $1-p$. Our goal is to exactly recover the k top-ranked items.

Underlying Model: When the set of items in $\mathcal{A}$ are being ranked, the probability of the resulted ranking being the order $v = (v_1, v_2, v_3)$ is denoted as $M_{v_1v_2v_3}$ or M_v. Note that here v is considered a vector and the sequence of the elements is important; we shall write $v \doteq \mathcal{A}$ if the items in the vector v are exactly those in the set $\mathcal{A}$. As a requirement of being a reasonable distribution, such M_v must satisfy $M_v \geq 0$ for any v with distinct elements, moreover,

$$\sum_{v \doteq \mathcal{A}} M_v = 1$$

We denote the collection of such valid probability assignment on M as $\mathcal{M}$.

2.2 Plackett-Luce Model

To be clear, our work makes no assumptions on the form of the ranking comparison probabilities. However, we will use Plackett-Luce Model, which is the most popular statistical model for ranking data, for simulation in Section 4. Our Plackett-Luce Model is restricted in case that

$|v| = 3$.

Plackett-Luce Model: The parameter space is $\Theta = \{\theta = (\theta_1, ..., \theta_m | \forall i, \theta_i \in (0, 1), \sum_{i=1}^{m} \theta_i = 1\}$. Given parameter $\theta \in \Theta$, the probability of any linear order $v = (v_1, v_2, v_3)$ is

$$\mathbf{Pr}(v|\theta) = \frac{\theta_{v_1}}{\theta_{v_1} + \theta_{v_2} + \theta_{v_3}} \times \frac{\theta_{v_2}}{\theta_{v_2} + \theta_{v_3}} \times \frac{\theta_{v_3}}{\theta_{v_3}}$$

In a Plackett-Luce model, each item i is associated with a positive parameter θ_i which represents the "quality" of the it. The greater the quality, the chance the item will be ranked at a higher position in comparison.

2.3 Generalized Copeland Algorithm

We addressed an algorithm to exactly recovering the k top-ranked items given noisy observations. The mechanism we consider to produce the ranking from noisy observations is based on scores: in a competition among the items in $\mathcal{A}$ where $|\mathcal{A}| = 3$, the i^{th} position item will receive score $\beta_{m,i}$ and every integer $l \in [r]$, let $Y^l_{a,\ \mathcal{A}}$ represents the l^{th} ranking result of item a, defined as

$$Y^l_{a,\ \mathcal{A}^-} = \begin{cases} \beta_1 & \text{if a ranks first} \\ \beta_2 & \text{if a ranks second} \\ \beta_3 & \text{if a ranks third} \end{cases}$$

where we assume $\beta_1 \geq \beta_2 \geq \beta_3 \geq 0$.

for $i \in [n]$ the quantity

$$N_i = \sum_{l \in [r]} \sum_{\mathcal{A} \subseteq [n] \backslash \{a\}} Y^l_{a,\ \mathcal{A}^-}$$

corresponds to the score of item i wins overall partial ranking, For each integer k, the vector $\{N_i\}_{i=1}^{n}$ defines a $k-$ sized subset

$$\widetilde{S}_k = \{i \in [n] \mid N_i \text{ is among the } k \text{ highest over } \{N_i\}_{i=1}^{n}\}$$

3. MAIN RESULT

In this proposal, we will propose a theorem of the optimality of the algorithm we addressed in section 2.2. The analysis consists of 2 parts: the forward direction (section 3.2), which tells us under what condition, the accuracy of results is guaranteed, and the converse direction (section 3.3), which provides the lower bound of the algorithm. I will discuss our intention of proof respectively, but before that, I want to mention the measurements of the degree of difficulty to separate the top k^{th} items d from the remaining, which will be used to describe the very condition under which the result is guaranteed to be accurately ranked. This is required by using a probabilistic way to prove.

3.1 Thresholds for Exact Recovery of the Top k Items

Let us use (k) and $(k+1)$ to denote the indices of the items that are ranked k^{th} and $(k+1)^{th}$ respectively. With this notation, the $k - separation\ threshold$ Δ_k is given by

$$\Delta k := \tau_{(k)} - \tau_{(k+1)}$$

Where the τ_a is defined as

$$\tau_a = \frac{1}{n^2 \beta_1}(\sum_{j,k} \beta_1 M_{ajk} + \sum_{j,k} \beta_2 M_{jak} + \sum_{j,k} \beta_3 M_{jka})$$

And the exact form of τ will be discussed in the remaining thesis.

3.2 Forward Direction

First define the following set of partial ranking probabilities:

$$\mathcal{F}_k(\alpha) = \{M \in \mathcal{M} : \Delta_k \geq \alpha \sqrt{\frac{log\ n}{rpn^2}}\}$$

Theorem 1: For any $\alpha > 0$, the probability of choosing incorrect top-k items using the score-based method for any items with $M \in \mathcal{F}_k(\alpha)$ is upper-bounded as

$$\sup_{M \in \mathcal{F}_k(\alpha)} \mathbb{P}_M[\hat{S}_k \neq S_k^*] \leq n^{-\frac{\alpha^2}{2\beta_1^2}}$$

the forward part intends to expound that as long as the top k^{th} items are easy enough to be separated from the remaining, which puts a constrain on the underlying model, the N_i computed by the algorithm is credible enough to discern top k items of the $[n]$.

Though we all believe that the samples will be finally follows the distribution of underlying model, one of the so-called "central inequalities" will tell us how fast the sampling will converge, which gives us the upper bound of our algorithm based on the constraints put on the underlying model by the degree of separation.

The process of our proof will follow the following cliche: finding the corresponding random variable of the underlying distribution, centralizing it, then bound its expectation and variance by $threshold\ \Delta k$, then wrap up all things using Berstain inequality. We would have the final proof.

3.3 Converse Direction

Theorem 2:Suppose that $n \geq 7$ and $p \geq \frac{logn}{36n^2r}$, Then for any $\alpha \leq \frac{(\beta_1-\beta_3)(n-k-1)}{10n\beta_1}$ the error probability of any estimator $\hat{S}_k$:

$$\sup_{M \in \mathcal{F}_{||}(\alpha)} \mathbb{P}_M[\hat{S}_k \neq S_k^*] \geq \frac{3}{10}$$

The intention of proving converse part is to prove that if the top k^{th} items are very hard to be separated from the remaining, then it is very hard to tell the underlying model from another, thus provide a lower bound for any estimator, which can be found by fano's inequality.

This part is more difficult since it requires building 2 distribution while maintaining the considerably simple relationship between the Δk and the KL divergence. We will do some mathematical simplification here to make the proof easier.

4. EXPERIMENT AND SIMULATION

We begin with simulations using generated data with observation probability $p = 1$, which means all comparison is observed.

4.1 $(\Delta_k, \overline{d})$ Plot

We will use this so-called $(\Delta_k, \overline{d})$ plot to represent the result. We generate each point of this plot by:

Firstly, Given the scale parameter (p, α, n, k) we can then compute the threshold of Δ_k^*

$$\Delta_k = \alpha\sqrt{\frac{log\ n}{rpn^2}}$$

Suppose we have assigned a preference vector θ carefully,

$$\theta = (\theta_1, \theta_2, \theta_3, ..., \theta_n)$$

then we use C to denote the collection of all possible rank τ_i

$$C = (\tau_1, \tau_2, \tau_3, ...)$$

Using PL model based on θ, we could compute an associate probability $P(\tau = \tau_i) = \xi_i$, thus we simulate the underlying model U controlled by parameter vector p

$$U = ((\tau_1, \xi_1), (\tau_2, \xi_2), ...)$$

Given the scale (n, m, k) we can then compute the threshold of Δ_k, then we will pick several U_i and V_j such that

$$U_i \in \mathcal{F}_k(\alpha)\ while\ V_j \notin \mathcal{F}_k(\alpha)$$

by assign corresponding θ carefully

Then sampling with these U_i and V_j, let $Y_{\mathcal{A}}$ denote the outcome of once sampling, we generate

$$Y_{i,\ \mathcal{A}} \sim U_i, \text{and } Y_{j,\ \mathcal{A}} \sim V_j$$

Run algorithm mentioned on section 2.3 over $Y_{i,\ \mathcal{A}},\ Y_{j,\ \mathcal{A}}$ to get $\widetilde{S}_{k,\ Y_{i,\ \mathcal{A}}}$ and $\widetilde{S}_{k,\ Y_{j,\ \mathcal{A}}}$, compare them with

$$S_k^* = \underset{top\ k}{argsort}(p)$$

here S_k^* is assumed to be true underlying ranking. For each θ assignment, we apply the algorithm, and then compute the average of d. which is

$$|d| = |\widetilde{S}_{k,\ Y_{i,\ \mathcal{A}}} - S_k^*|$$

Here, d can be seen as $l-1$ norm of the difference between 2 vectors. Thus, for each θ, we can have a $(\Delta_k, \overline{d})$ pair.

4.2 Result

In this paper, I have done 2 simulations, one is for $n = 20, k = 4$, another is $n = 50, k = 4$

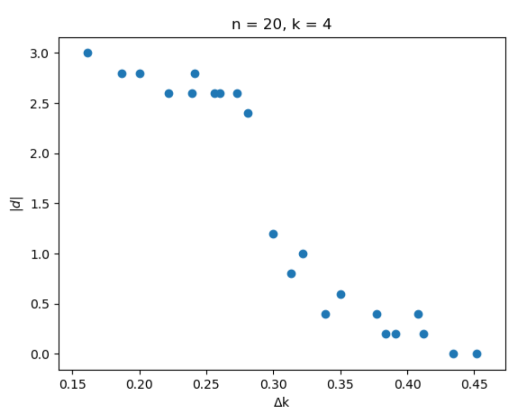

Figure 4.1: Simulation 1. $n = 20, k = 4$

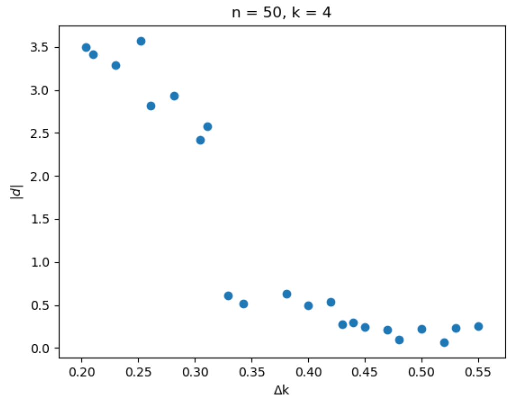

Figure 4.2: Simulation 2. $n = 50, k = 4$

Figure 4.1 shows that, for $n = 20, k = 4$, the threshold is in $[0.22, 0.27]$. Figure 4.2 shows that, for $n = 50, k = 4$, the threshold is in $[0.29, 0.31]$.

Both Simulation show that if the $\Delta_k \leq \Delta_k^*$, then the algorithm's outcome will be different from the underlying model very likely, while if $\Delta_k \geq \Delta_k^*$, then the algorithm's result will probably reflect the true underlying model. Though the behavior around the threshold is still unknown.

We can also see that along with the enlarge of data scale, the ability of algorithms to discrete true ranking improves.

5. PROOF

We now turn to the proofs of our main results. We continue to use the notation $[i]$ to denote the set $\{1, ..., i\}$ for any integer $i \geq 1$.

Our upper bound is based on Bernstein Inequality, which is shown in the Forward Part. While is lower bound is derived from Fano's inequality.

Lemma 1: *Bernstein Inequality*. Let $X_1, ..., X_n$ be independent zero-mean random variables, Suppose that for all $i, |X_i| \leq M$, Then for all positive t, we have that

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i > t\right) \leq exp\left(-\frac{\frac{1}{2}t^2}{\sum_i \mathbb{E}[X_i^2] + \frac{1}{3}Mt}\right)$$

Lemma 2: *Fano's Inequality.* Fix some integer $L \geq 2$, fix some collection of distributions $\{\mathbb{P}^{\mathbb{1}}, ..., \mathbb{P}^{\mathbb{L}}\}$, adn then drawing $Y \sim \mathbb{P}^{\mathbb{A}}$. Using $\mathcal{Y}$ to denote the sample space associated with the observation Y, Fano's inequality asserts that any test function $\phi : \mathcal{Y} \to [L]$, for this problem has error probability lower bounded as

$$\mathbb{P}[\phi(Y) \neq A] \geq 1 - \frac{I(Y; A) + log\ 2}{log\ L}$$

Where $I(Y; A)$ denotes the mutual information between Y and A. A standard convexity argument for the mutual information yields the weaker bound

$$\mathbb{P}[\phi(Y) \neq A] \geq 1 - \frac{\max\limits_{a,\ b \in [L]} D_{KL}(\mathbb{P}^a || \mathbb{P}^b) + log\ 2}{log\ L}$$

5.1 Forward Part

First define the following set of partial ranking probabilities:

$$\mathcal{F}_k(\alpha) = \{M \in \mathcal{M} : \Delta_k \geq \alpha \sqrt{\frac{log\ n}{rpn^2}}\} \tag{1}$$

recall the definition of Δ_k is

$$\Delta_k \triangleq \tau_a - \tau_b = \frac{1}{n^2}\left(\sum_{jk}\beta_1(M_{ajk} - M_{bjk}) + \sum_{jk}\beta_1(M_{jak} - M_{jbk}) + \sum_{jk}\beta_1(M_{jka} - M_{jkb})\right) \tag{2}$$

Theorem 1 For any $\alpha > 0$, the probability of choosing incorrect top-k items using the score-based method for any items with $M \in \mathcal{F}_k(\alpha)$ is upper-bounded as

$$\underset{M\in\mathcal{F}_{||}(\alpha)}{sup} \mathbb{P}_M[\hat{S}_k \neq S_k^*] \leq n^{-\frac{\alpha^2}{2\beta_1^2}}$$

Proof. Consider any item $a \in S^*$ and $b \in [n]\ S^*$, then define the event that $W_b > W_a$ as $\epsilon_{ba} = W_b > W_a$, i.e.,

$$\mathbf{Pr}(\epsilon_{ba}) = \mathbf{Pr}(W_b - W_a > 0). \tag{3}$$

We here use $X^{(l)}_{a,\ \mathcal{A}^-}$ to denote the score item a receives in the l - th round in the competition among the items in $\mathcal{A} = a\ \cup \mathcal{A}^- \subseteq [n]$, thus it is distributed as follows

$$\mathbf{Pr}(X^{(l)}_{a,\ \mathcal{A}^-} = \beta) = \begin{cases} pR^t_{a,\ \mathcal{A}^-} & \beta = \beta_t\ , t = 1, 2, 3 \\ 1 - p & \beta = 0 \end{cases}$$

Where $R^t_{a,\ \mathcal{A}^-}$ is the probability that item - a ranks at the t-th position in the set $\{a\} \bigcup \mathcal{A}^-$, and thus

$$\begin{cases} R^1_{a,\ \mathcal{A}^-} = M_{ajk} + M_{akj} \\ R^2_{a,\ \mathcal{A}^-} = M_{ajk} + M_{akj} \\ R^3_{a,\ \mathcal{A}^-} = M_{ajk} + M_{akj} \end{cases} \tag{4}$$

After a total of r rounds of comparison, the score item-a receives is thus given as

$$W_a = \sum_{l\in[r]} \sum_{\mathcal{A}^- \subseteq [n]\backslash\{a\}} X^{(l)}_{a,\ \mathcal{A}^-}$$

Thus, from (3)

$$W_b - W_a = \sum_{l\in[r]} \sum_{\mathcal{A}^- \subseteq [n]\backslash\{b\}} X^{(l)}_{b,\ \mathcal{A}^-} - \sum_{l\in[r]} \sum_{\mathcal{A}^- \subseteq [n]\backslash\{a\}} X^{(l)}_{a,\ \mathcal{A}^-} \geq 0 \tag{5}$$

To utilize Lemma 1. we first centralized $X^{(l)}_{a,\ \mathcal{A}^-}, X^{(l)}_{b,\ \mathcal{A}^-}$ for upcoming operation:

$$\overline{X}^{(l)}_{b,\ \mathcal{A}^-} \triangleq X^{(l)}_{b,\ \mathcal{A}^-} - \mathbb{E}(\overline{X}^{(l)}_{b,\ \mathcal{A}^-}) = X^{(l)}_{b,\ \mathcal{A}^-} - p\sum_{t=1}^{3} \beta_t R^t_{b,\ \mathcal{A}^-},$$

similarly,

$$\overline{X}^{(l)}_{a,\ \mathcal{A}^-} \triangleq X^{(l)}_{a,\ \mathcal{A}^-} - \mathbb{E}(\overline{X}^{(l)}_{a,\ \mathcal{A}^-}) = X^{(l)}_{a,\ \mathcal{A}^-} - p\sum_{t=1}^{3} \beta_t R^t_{a,\ \mathcal{A}^-},$$

and second, define and centralize cross-score $\overline{X}^{(l)}_{\{a,b\},\ \mathcal{A}^{--}}$:

$$\begin{aligned}\overline{X}^{(l)}_{\{a,b\},\ \mathcal{A}^{--}} &\triangleq X^{(l)}_{a,\ \{b,\ \mathcal{A}^{--}\}} - X^{(l)}_{b,\ \{a,\ \mathcal{A}^{--}\}} - \mathbb{E}(X^{(l)}_{a,\ \{b,\ \mathcal{A}^{--}\}}) + \mathbb{E}(X^{(l)}_{b,\ \{a,\ \mathcal{A}^{--}\}}) \\ &= X^{(l)}_{a,\ \{b,\ \mathcal{A}^{--}\}} - X^{(l)}_{b,\ \{a,\ \mathcal{A}^{--}\}} - p\left(\sum_{t=1}^{3} \beta_t R^t_{a,\ \{b,\ \mathcal{A}^{--}\}} - \sum_{t=1}^{3} \beta_t R^t_{b,\ \{a,\ \mathcal{A}^{--}\}}\right)\end{aligned}$$

The $X^{(l)}_{a,\ \mathcal{A}^-}, X^{(l)}_{b,\ \mathcal{A}^-}, \overline{X}^{(l)}_{\{a,b\},\ \mathcal{A}^{-}}$ is now all zero-mean, and mutually independent, which is the prerequisite to apply the Lemma 1. Bernstein Inequality:

Given (4) ,

$$\begin{aligned}&\sum_{\mathcal{A}^{--}\subseteq[\backslash]\backslash\{\dashv,\lfloor\}} \left(p\sum_{t=1}^{3} \beta_t R^t_{b,\ \mathcal{A}^{--}} - p\sum_{t=1}^{3} \beta_t R^t_{a,\ \mathcal{A}^{--}}\right) \\ &+ \sum_{\mathcal{A}^{--}\subseteq[\backslash]\backslash\{\dashv,\lfloor\}} p\left(\sum_{t=1}^{3} \beta_t R^t_{a,\ \{b,\ \mathcal{A}^{--}\}} - \sum_{t=1}^{3} \beta_t R^t_{b,\ \{a,\ \mathcal{A}^{--}\}}\right) \\ &= rp\sum_{jk} \beta_1(M_{bjk} - M_{ajk}) + \beta_2(M_{jbk} - M_{jak}) + \beta_3(M_{jkb} - M_{jka}) \geq rpn^2\Delta_k\end{aligned}$$

Follows (5) that

$$\mathbb{P}(\epsilon_{ba})$$
$$= \mathbb{P}\left(\sum_{l\in[r]}\left(\sum_{\mathcal{A}^{--}\subseteq[n]\setminus\{a,\, b\}}\left(\overline{X}^{(l)}_{b,\, \mathcal{A}^{-}} - \overline{X}^{(l)}_{a,\, \mathcal{A}^{-}}\right) + \sum_{\mathcal{A}^{--}\subseteq[n]\setminus\{a,\, b\}}\overline{X}^{(l)}_{\{a,b\},\, \mathcal{A}^{--}}\right) \geq rpn^2\Delta_k\right) \tag{6}$$

Another relation we can derived from (4) is

$$\sum_{t}^{3} R^t_{a,\, \mathcal{A}^-} = 1$$

for any $a,\ \mathcal{A}^-$.

thus,

$$\mathbb{E}(\overline{X}^{(l)}_{b,\, \mathcal{A}^-})^2 = \sum_{t=1}^{3}\beta_t^2 pR^t_{b,\, \mathcal{A}^-} - \left(\sum_{t=1}^{3}\beta_1^2 pR^t_{b,\, \mathcal{A}^-}\right)^2 \leq \sum_{t=1}^{3}\beta_t^2 pR^t_{b,\, \mathcal{A}^-}$$

Similarly,

$$\mathbb{E}(\overline{X}^{(l)}_{a,\, \mathcal{A}^-})^2 = \sum_{t=1}^{3}\beta_t^2 pR^t_{a,\, \mathcal{A}^-} - \left(\sum_{t=1}^{3}\beta_1^2 pR^t_{a,\, \mathcal{A}^-}\right)^2 \leq \sum_{t=1}^{3}\beta_t^2 pR^t_{a,\, \mathcal{A}^-}$$

And for $\overline{X}^{(l)}_{\{a,b\},\, \mathcal{A}^{--}}$

$$\begin{aligned}\mathbb{E}(\overline{X}^{(l)}_{\{a,b\},\, \mathcal{A}^{--}})^2 &= \mathbb{E}(X^{(l)}_{a,\, \{b,\, \mathcal{A}^{--}\}} - X^{(l)}_{b,\, \{a,\, \mathcal{A}^{--}\}})^2 - (\mathbb{E}X^{(l)}_{a,\, \{b,\, \mathcal{A}^{--}\}} - \mathbb{E}X^{(l)}_{b,\, \{a,\, \mathcal{A}^{--}\}})^2 \\ &\leq \sum_{t=1}^{3} p\beta_t R^t_{b,\{a,\, \mathcal{A}^{--}\}} + p\beta_t R^t_{a,\{b,\, \mathcal{A}^{--}\}}\end{aligned}$$

Thus,

$$
\begin{aligned}
&\sum_{l\in[r]}\left[\sum_{\mathcal{A}^{--}\subseteq[n]\backslash\{a,\,b\}}\left(\overline{X}_{b,\,\mathcal{A}^{-}}^{(l)}\right)^2+\sum_{\mathcal{A}^{--}\subseteq[n]\backslash\{a,\,b\}}\left(\overline{X}_{a,\,\mathcal{A}^{-}}^{(l)}\right)^2+\sum_{\mathcal{A}^{--}\subseteq[n]\backslash\{a,\,b\}}\left(\overline{X}_{\{a,b\},\,\mathcal{A}^{--}}^{(l)}\right)^2\right]\\
&\leq rp\sum_{\mathcal{A}^{--}\subseteq[n]\backslash\{a,\,b\}}\sum_{t=1}^{3}\left[p\beta_t R^t_{b,\{a,\,\mathcal{A}^{--}\}}+p\beta_t R^t_{a,\{b,\,\mathcal{A}^{--}\}}\right]\\
&\leq rp\left[(\beta_1^2-\beta_3^2)\sum_{\mathcal{A}^{--}\subseteq[n]\backslash\{a,\,b\}}R^1_{b,\{a,\,\mathcal{A}^{--}\}}+(\beta_2^2-\beta_3^2)\sum_{\mathcal{A}^{--}\subseteq[n]\backslash\{a,\,b\}}R^2_{b,\{a,\,\mathcal{A}^{--}\}}+\beta_3^2\frac{n^2}{2}\right]\\
&+rp\left[(\beta_1^2-\beta_3^2)\sum_{\mathcal{A}^{--}\subseteq[n]\backslash\{a,\,b\}}R^1_{a,\{b,\,\mathcal{A}^{--}\}}+(\beta_2^2-\beta_3^2)\sum_{\mathcal{A}^{--}\subseteq[n]\backslash\{a,\,b\}}R^2_{a,\{b,\,\mathcal{A}^{--}\}}+\beta_3^2\frac{n^2}{2}\right]
\end{aligned}
\tag{7}
$$

Since

$$
\tau_a-\tau_b\geq\Delta_k
$$

It follows (7) that

$$
\begin{aligned}
&\sum_{l\in[r]}\left[\sum_{\mathcal{A}^{--}\subseteq[n]\backslash\{a,\,b\}}\left(\overline{X}_{b,\,\mathcal{A}^{-}}^{(l)}\right)^2+\sum_{\mathcal{A}^{--}\subseteq[n]\backslash\{a,\,b\}}\left(\overline{X}_{a,\,\mathcal{A}^{-}}^{(l)}\right)^2+\sum_{\mathcal{A}^{--}\subseteq[n]\backslash\{a,\,b\}}\left(\overline{X}_{\{a,b\},\,\mathcal{A}^{--}}^{(l)}\right)^2\right]\\
&\leq 2rp(\beta_1^2-\beta_3^2)\sum_{\mathcal{A}^{--}\subseteq[n]\backslash\{a,\,b\}}R^1_{a,\,\{a,\,\mathcal{A}^{--}\}}\\
&+rp(\beta_1+\beta_2+2\beta_3)(\beta_2-\beta_3)\sum_{\mathcal{A}^{--}\subseteq[n]\backslash\{a,\,b\}}R^2_{a,\,\{a,\,\mathcal{A}^{--}\}}+rpn^2\beta_3^2\\
&-rp(\beta_1-\beta_2)(\beta_2-\beta_3)\sum_{\mathcal{A}^{--}\subseteq[n]\backslash\{a,\,b\}}R^2_{a,\,\{b,\,\mathcal{A}^{--}\}}-(\beta_1+\beta_3)rpn^2\Delta_k\\
&\leq 2rpn^2\beta_1^2-(\beta_1+\beta_3)rpn^2\Delta_k
\end{aligned}
\tag{8}
$$

Plug in (8) and the (6) to Lemma 1:

$$
\begin{aligned}
\mathbf{Pr}(\epsilon_{ba})=\mathbf{Pr}(W_b-W_a>0)&\leq exp\left(-\frac{\frac{1}{2}\alpha^2(rpn^2)^2\Delta_k^2}{rpn^2\beta_1^2-rpn^2(\beta_1+\beta_3)\Delta_k+\frac{2}{3}rpn^2\beta_1}\right)\\
&\leq exp(-\frac{\alpha\ log\ n}{4})=n^{-\alpha^2/4}
\end{aligned}
$$

recall the α is the constant in (1).

5.2 Converse Part

For each $a \in \{k, ..., n\}$, denote the $k-$sized subset $S^*[a] = \{1, 2, ..., k-1\} \bigcup \{a\}$, $v = (v_1, v_2, v_3)$

$$M_v \triangleq \begin{cases} \frac{1}{6} + \delta & \text{if } v_1 \in S^*[a] \text{ and } v_2, v_3 \notin S^*[a] \\ \frac{1}{6} - \delta & \text{if } v_1, v_2 \in S^*[a] \text{ and } v_3 \notin S^*[a] \\ \frac{1}{6} & \text{other cases} \end{cases}$$

Thus, for the threshold Δ_k

$$\Delta_k = \frac{(\beta_1 - \beta_3)(n-k-1)\delta}{n\beta_1}$$

Theorem 2. Suppose that $n \geq 7$ and $p \geq \frac{logn}{36n^2r}$, Then for any $\alpha \leq \frac{(\beta_1-\beta_3)(n-k-1)}{10n\beta_1}$ the error probability of any estimator $\hat{S}_k$:

$$\sup_{M \in \mathcal{F}_{||}(\alpha)} \mathbb{P}_M[\hat{S}_k \neq S^*_k] \geq \frac{3}{10}$$

Proof: we prove Theorem 2 by introducing the Proposition 1 first,

Proposition 1. For any distnact $a,\ b \in \{1, 2, ..., n\}$, we have

$$D_{KL}(\mathbb{P}^{\supset}||\mathbb{P}^b) \leq \frac{(n-k-1)(n+k-\frac{1}{3})pr\delta^2}{2(\frac{1}{36} - \delta^2)}$$

Then apply the Lemma 2. (Fano's inequality), we have,

$$\mathbb{P}[\phi \neq A] \geq 1 - \frac{+log\ 2}{log(n-k+1)} \geq \frac{3}{10}$$

if given that $\delta \leq \frac{1}{10}\sqrt{\frac{logn}{npr}}$, $p \geq \frac{logn}{36n^2r}$, and $n \geq 7$

the next question is the proof of Proposition 1.

Proof of Proposition 1:

for $v = (v_1, v_2, v_3) \doteq \mathcal{A}$

$$D_{KL}(\mathbb{P}^a||\mathbb{P}^b) = rp\sum_{l\in[r]}\sum_{\mathcal{A}\subseteq[n]} D_{KL}(\mathbb{P}^a(V_{\mathcal{A}})||\mathbb{P}^b(V_{\mathcal{A}}))$$

where $V_{\mathcal{A}}$ represent the outcome of the comparison among the elements in $\mathcal{A}$

Case 1. If $v_1 = a,\ v_2, v_3 \in [n]\backslash\{a, b\})$, then clearly

$$D_{KL}(\mathbb{P}^a(V_{\mathcal{A}})||\mathbb{P}^b(V_{\mathcal{A}})) = 0$$

Because the distribution of $V_{\mathcal{A}}$ is identical across $\mathbb{P}^a$ and $\mathbb{P}^b$

Case2.. If $v_1 = a,\ v_2, v_3 \in [n]\backslash([k-1] \cup \{a, b\})$

$$D_{KL}(\mathbb{P}^a(V_{\mathcal{A}})||\mathbb{P}^b(V_{\mathcal{A}})) = 2(\frac{1}{6}+\delta)plog(\frac{\frac{1}{6}+\delta}{\frac{1}{6}}) + 2(\frac{1}{6}-\delta)p\log(\frac{\frac{1}{6}-\delta}{\frac{1}{6}}) \leq 24\delta^2 \leq \frac{\frac{2}{3}\delta^2 p}{(\frac{1}{6})^2-\delta^2}$$

Case 3. If $v_1 = a,\ v_2 \in [n]\backslash([k-1] \cup \{a, b\}), v_3 \in [k-1]$

$$D_{KL}(\mathbb{P}^a(V_{\mathcal{A}})||\mathbb{P}^b(V_{\mathcal{A}})) = 2(\frac{1}{6})plog(\frac{\frac{1}{6}}{\frac{1}{6}+\delta}) + 2(\frac{1}{6})plog(\frac{\frac{1}{6}}{\frac{1}{6}-\delta}) \leq \frac{\frac{1}{3}\delta^2 p}{(\frac{1}{6})^2-\delta^2}$$

Case 4. If $v_1 = b,\ v_2, v_3 \in [n]\backslash([k-1] \cup \{a, b\})$

$$D_{KL}(\mathbb{P}^a(V_{\mathcal{A}})||\mathbb{P}^b(V_{\mathcal{A}})) = 2(\frac{1}{6})plog(\frac{\frac{1}{6}}{\frac{1}{6}+\delta}) + 2(\frac{1}{6})plog(\frac{\frac{1}{6}}{\frac{1}{6}-\delta}) \leq \frac{\frac{1}{3}\delta^2 p}{(\frac{1}{6})^2-\delta^2}$$

Case 5. $v_1 = b,\ v_2 \in [n]\backslash([k-1] \cup \{a, b\}), v_3 \in [k-1]$

$$D_{KL}(\mathbb{P}^a(V_{\mathcal{A}})||\mathbb{P}^b(V_{\mathcal{A}})) = 2(\frac{1}{6}+\delta)plog(\frac{\frac{1}{6}+\delta}{\frac{1}{6}}) + 2(\frac{1}{6}-\delta)p\log(\frac{\frac{1}{6}-\delta}{\frac{1}{6}}) \leq 24\delta^2 \leq \frac{\frac{2}{3}\delta^2 p}{(\frac{1}{6})^2-\delta^2}$$

Case 6. $v_1 = a,\ v_2 = b,\ v_3 \in [n]\backslash([k-1] \cup \{a, b\})$

$$D_{KL}(\mathbb{P}^a(V_{\mathcal{A}})||\mathbb{P}^b(V_{\mathcal{A}})) = (\frac{1}{6}+\delta)plog(\frac{\frac{1}{6}+\delta}{\frac{1}{6}}) + (\frac{1}{6}-\delta)p\log(\frac{\frac{1}{6}-\delta}{\frac{1}{6}}) + \frac{1}{6}plog(\frac{\frac{1}{6}}{\frac{1}{6}+\delta}) + \frac{1}{6}plog(\frac{\frac{1}{6}}{\frac{1}{6}-\delta}) + (\frac{1}{6}+\delta)plog(\frac{\frac{1}{6}+\delta}{\frac{1}{6}-\delta}) + (\frac{1}{6}-\delta)plog(\frac{\frac{1}{6}-\delta}{\frac{1}{6}+\delta}) \leq \frac{\frac{11}{6}\delta^2 p}{(\frac{1}{6})^2-\delta^2}$$

Combine the bounds from all 6 cases, we can bound the KL divergence as below:

$$\begin{aligned} D_{KL}(\mathbb{P}^a(V_{\mathcal{A}})||\mathbb{P}^b(V_{\mathcal{A}})) &\leq \frac{(n-k-1)(n-k-2)pr\delta^2}{2((\frac{1}{6})^2-\delta^2)} + \frac{(n-k-1)(k-1)pr\delta^2}{(\frac{1}{6})^2-\delta^2} \\ &+ \frac{(n-k-1)\frac{11}{6}\delta^2 p}{(\frac{1}{6})^2-\delta^2} \\ &\leq \frac{(n-k-1)(n-k+\frac{5}{3})pr\delta^2}{2((\frac{1}{6})^2-\delta^2)} \end{aligned}$$

Thus, the Proposition 1. is proved.

6. SUMMARY

In summary, the key idea behind Copeland Counting Algorithm is that exact ranking can be learned from data with relatively high precision via simple calculation such as counting. Thus we generalized this counting algorithm from pairwise - comparison to triple - comparison case, which is a fairly original and useful try as we discussed in Chapter 2.

Through a similar proving framework, we provided the upper bound using Bernstein Inequality. Although the coping items made the independent variable much harder then pairwise cases, we still handled this problem and provided relatively tight bound at the end. Same as the lower bound, we found a balance between the tighter bound and simpler ways to describe the worst cases.

Moreover, we also tested the real performance of our ranking algorithm. Designed a measurement of ranking error, we are able to show that the triple case counting algorithm is working fairly fine on a large enough dataset, and the threshold of k-separation become more precise when a dataset is larger.

REFERENCES

[1] C. Kenyon-Mathieu and W. Schudy, "How to rank with few errors," *In Symposium on Theory of Computing (STOC)*, pp. 95–103, 2007.

[2] M. Braverman and E. Mossel, "Noisy sorting without resampling," *Proc. ACM- SIAM Symposium on Discrete Algorithms*, p. 268–276, 2008.

[3] B. Eriksson, "Learning to top-k search using pairwise comparisons," *Conference on Artificial Intelligence and Statistics*, 2013.

[4] D. Hunter, "Machine algorithms for generalized bradley-terry models," *Annals of Statistics*, p. 384–406, 2004.

[5] S. Negahban, S. Oh, and D. Shah, "Interactive ranking from pair-wise comparisons," *In Advances in Neural Information Processing Systems*, 2012.

[6] B. Hajek, S. Oh, and J. Xu, "Minimax-optimal inference from partial rankings," *In Advances in Neural Information Processing Systems*, 2014.

[7] H. Soufiani, D. Parkes, and L. Xia, "Computing parametric ranking models via rank-breaking," *International Conference on Machine Learning*, 2014.

[8] N. B. Shah, S. Balakrishnan, A. P. J. Bradley, K. Ramchandran, , and M. J. Wainwright, "Estimation from pairwise comparisons: Sharp minimax bounds with topology dependence," *Journal on Machine Learning Research*, 2016.

[9] S. Jagabathula and D. Shah, "Inferring rankings under constrained sensing," *In Advances in Neural Information Processing Systems*, 2008.

[10] I. Mitliagkas, A. Gopalan, C. Caramanis, and S. Vishwanath, "User rankings from comparisons: Learning permutations in high dimensions," *Allerton Conference on Communication, Control, and Computing*, 2011.

[11] A. Ammar and D. Shah, “Efficient rank aggregation using partial data,” *CM SIGMETRICS Performance Evaluation Review*, 2012.

[12] W. Ding, P. Ishwar, and V. Saligrama, “A topic modeling approach to ranking,” *In Conference on Artificial Intelligence and Statistics*, 2015.

[13] F. Wauthier, M. Jordan, and N. Jojic, “Efficient ranking from pairwise comparisons,” *International Conference on Machine Learning*, 2013.

[14] A. Rajkumar and S. Agarwal, “A statistical convergence perspective of algorithms for rank aggregation from pairwise data,” *International Conference on Machine Learning*, 2014.

[15] A. Rajkumar, S. Ghoshal, L.-H. Lim, and S. Agarwal, “Ranking from stochastic pairwise preferences: Recovering condorcet winners and tournament solution sets at the top.,” *International Conference on Machine Learning*, 2015.

[16] Y. Chen and C. Suh, “Spectral mle: Top-k rank aggregation from pairwise comparisons,” *International Conference on Machine Learning*, 2015.

[17] R. Bradley and M. Terry, “Rank analysis of incomplete block designs: I. the method of paired comparisons,” *Biometrika*, p. 324–345, 1952.